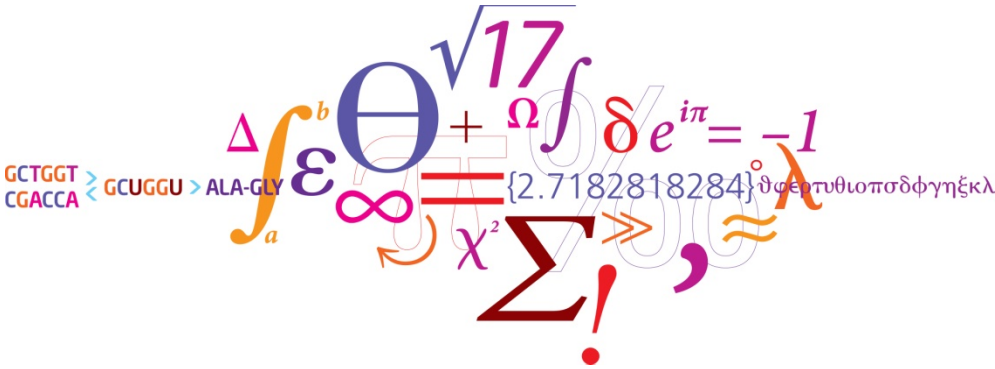
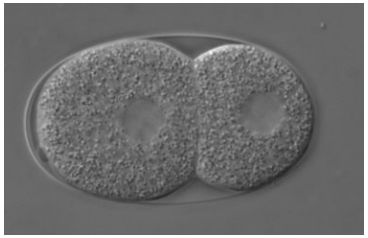
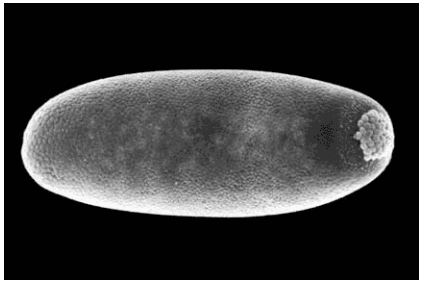
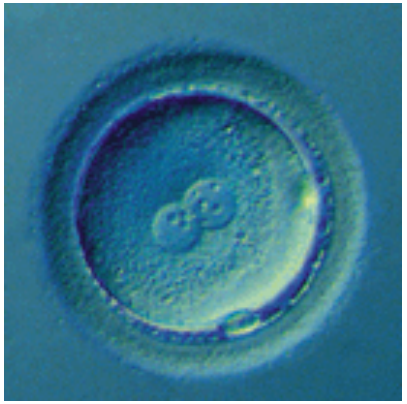
Measuring Protein-DNA interactions



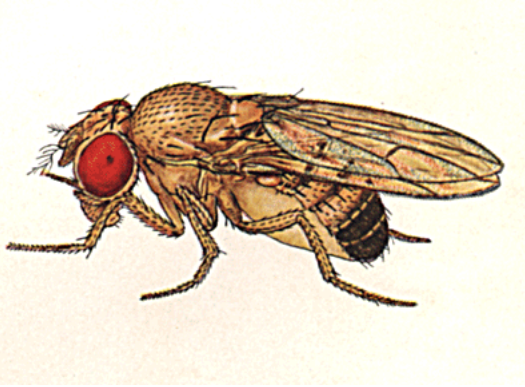
Center for Biological Sequence Analysis
Department of Systems Biology

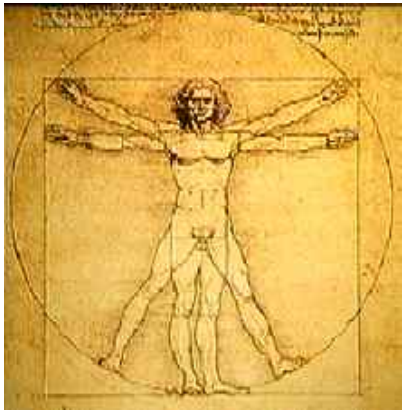
How is Biological Complexity Achieved?

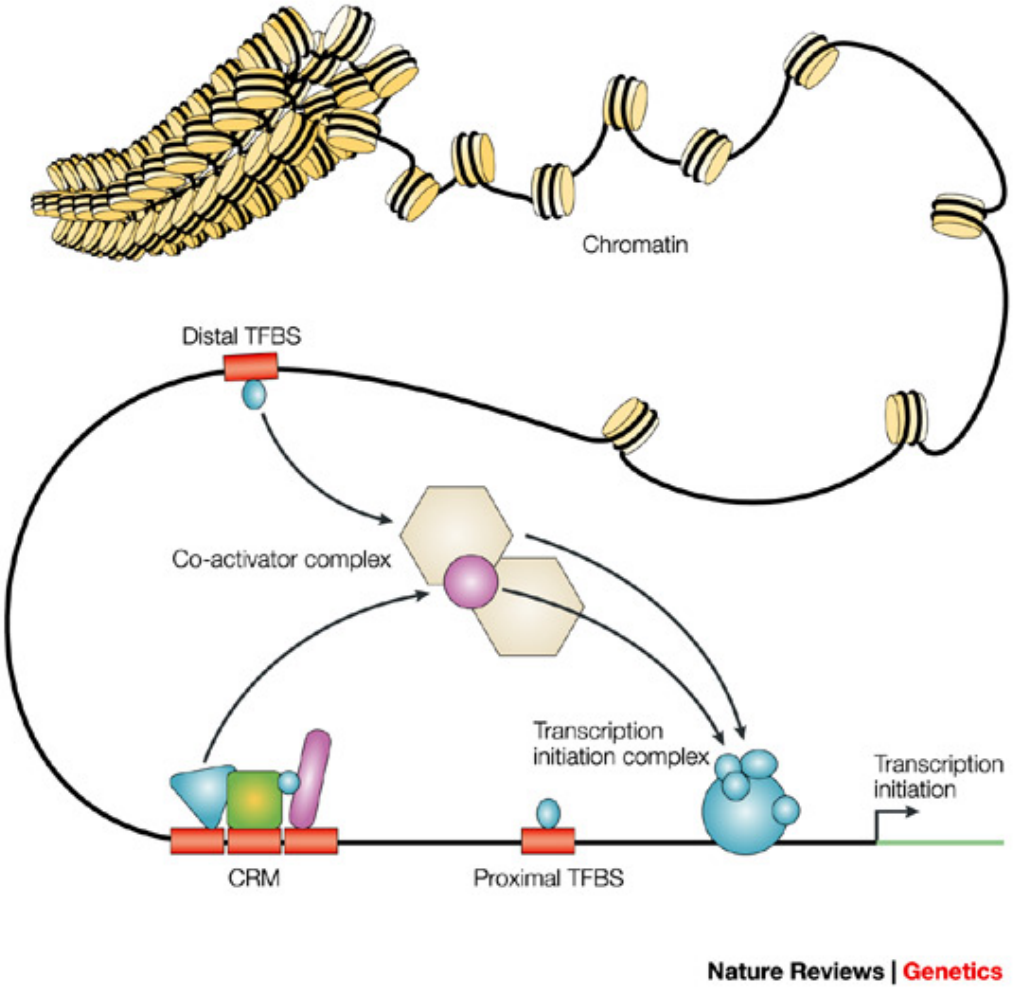


Mediated by Transcription Factors (TFs)



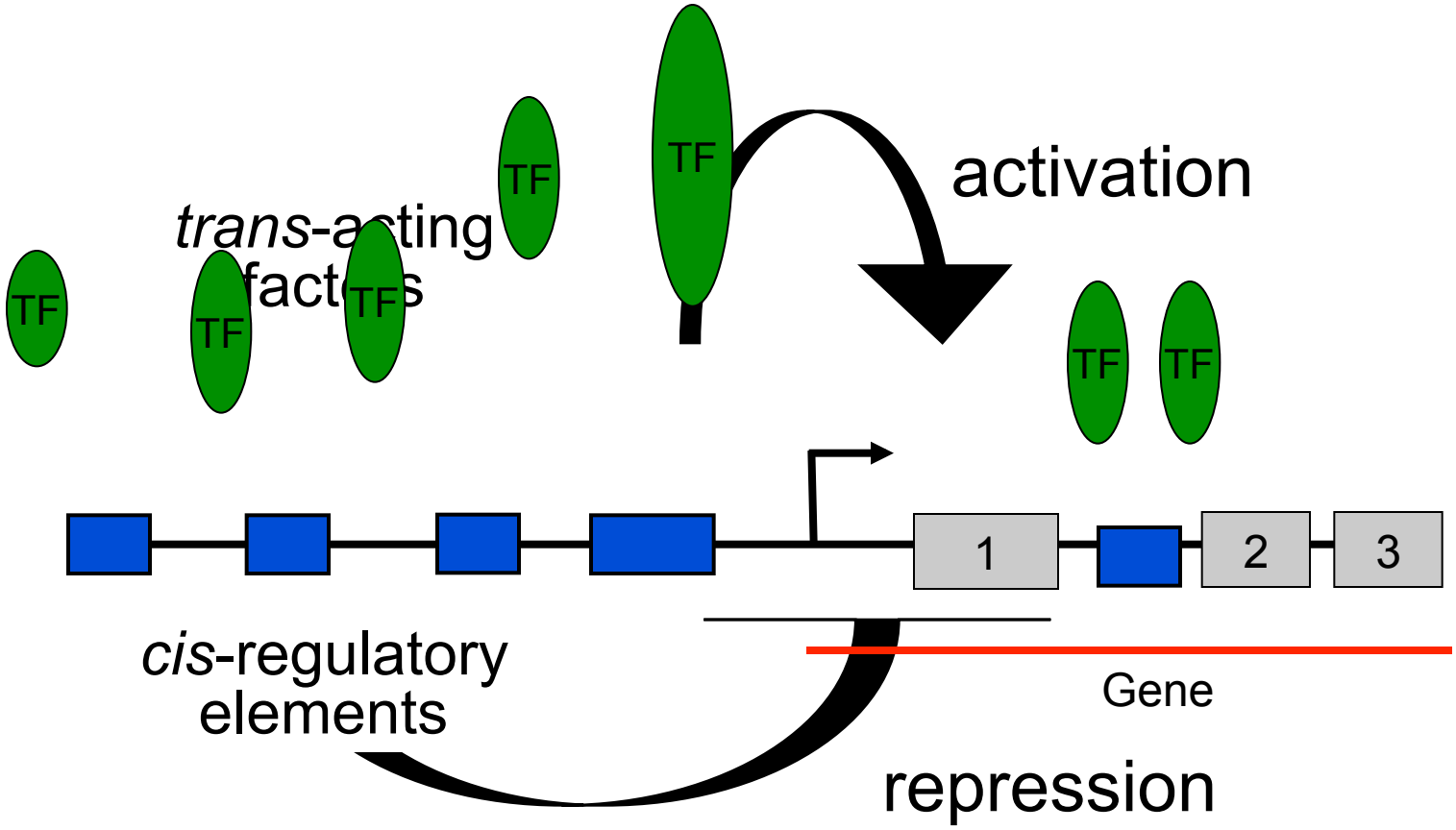


Transcription Factors are genetic switches



Wasserman & Sandelin Nature Reviews Genetics 2004

Regulation of Gene Expression by Transcription Factors



The big point is:

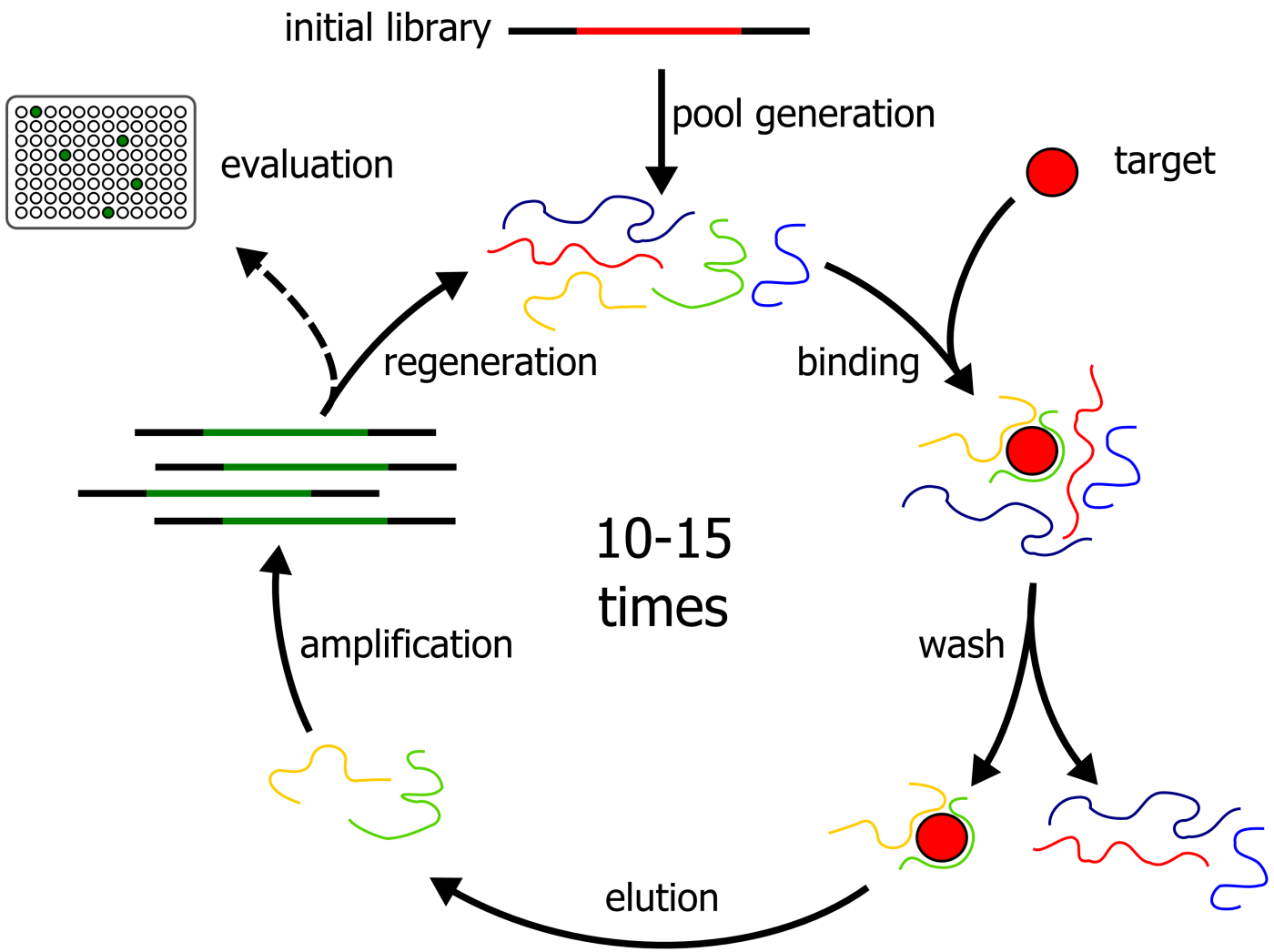
"…how these TFs orchestrate the expression of thousands of genes in a genome to create such a spectrum of biological diversity remains a mystery…"

Several methods have been developed in the last several years to study TF-DNA interactions and to understand the function of TFs.

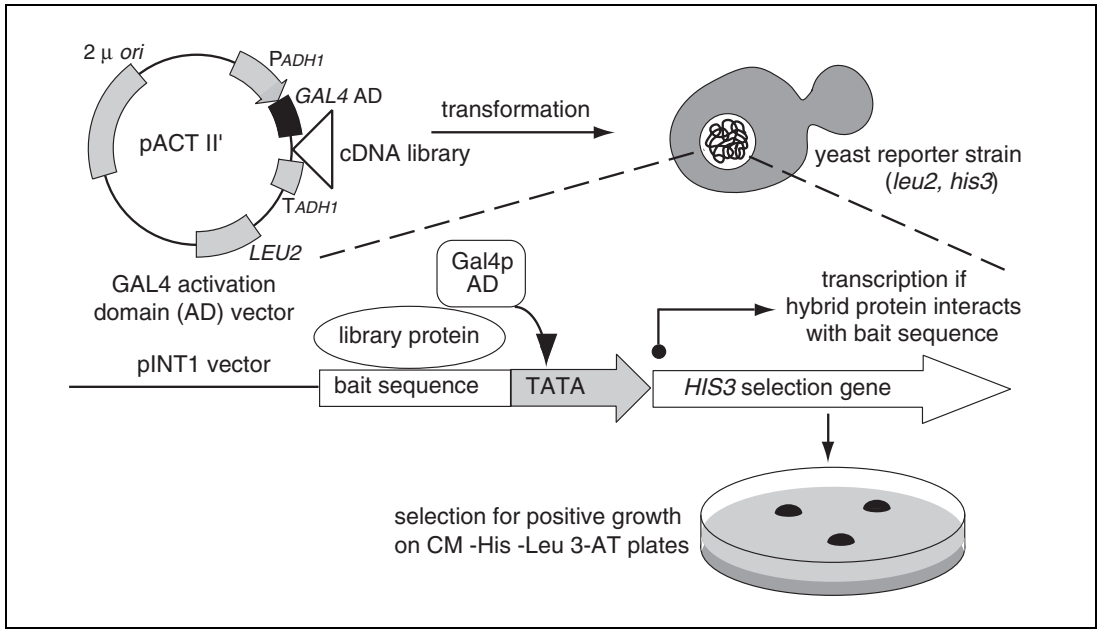
HTP Methods for studying TF-DNA interactions

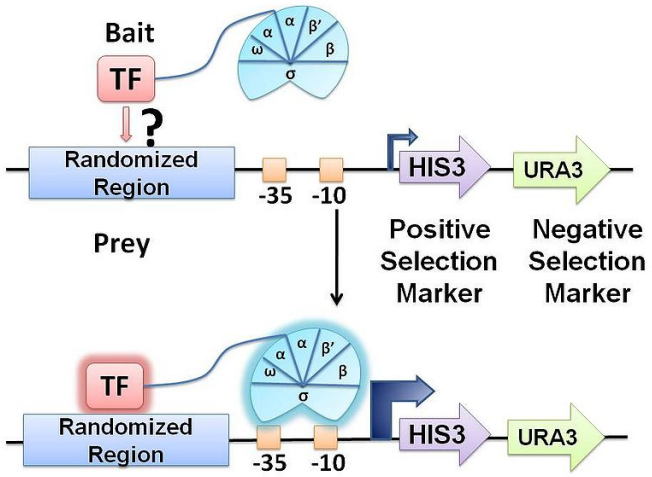
- Systematic Evolution of Ligands by Exponential Enrichment - SELEX (obsolete)
- Yeast-1-Hybrid (Y1H)
- Bacterial-1-Hybrid (B1H)
- Protein Binding Microarrays
- Chromatin Immunoprecipitation followed by chip (ChIP-Chip) or followed by Sequencing (ChIP-Seq)

SELEX


initial library
pool generation
target
evaluation
regeneration
binding
10-15
times
amplification
wash
elution

Y1H and B1H

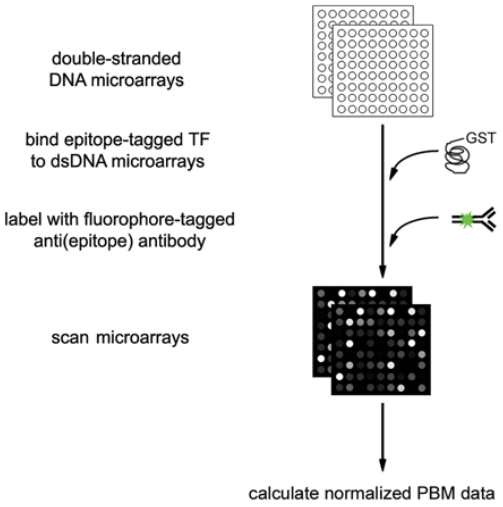




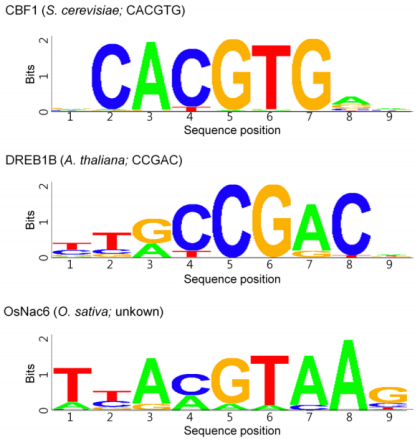
- Both are modifications of the Yeast-2-hybrid system.
- In Y1H, we screen several TFs (prey) against a fixed promoter (bait).
- In B1H, we screen milions of promoters (prey) against a fixed TF (bait)
- Surviving colonies are grown, the DNAs are sequenced (NGS) using and TF-specific DNA sequences retrieved for further analysis.

Protein Binding Microarrays

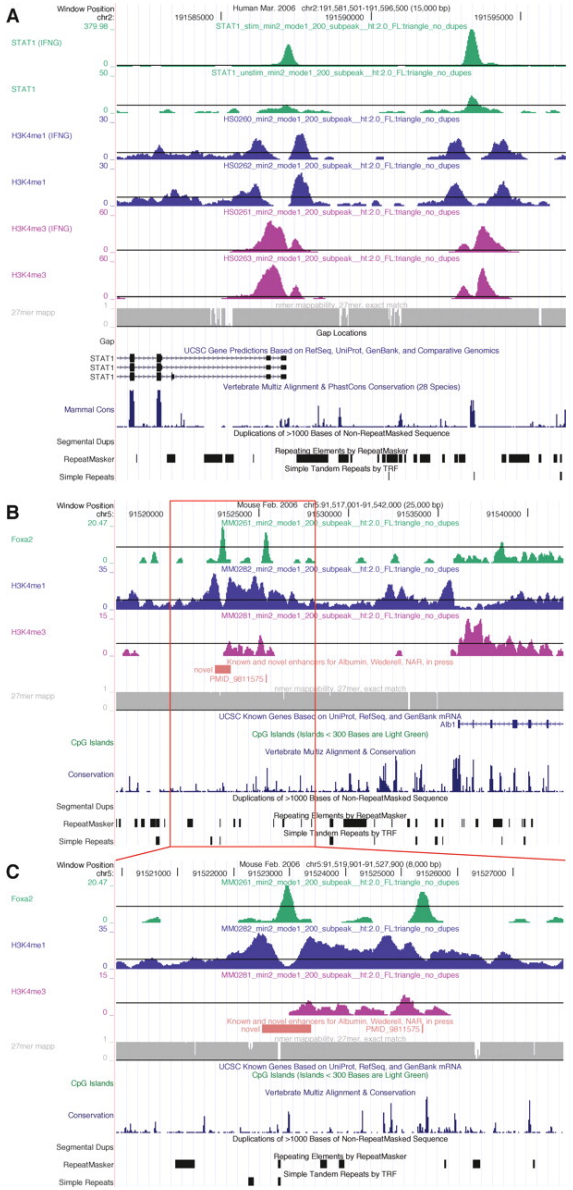
1



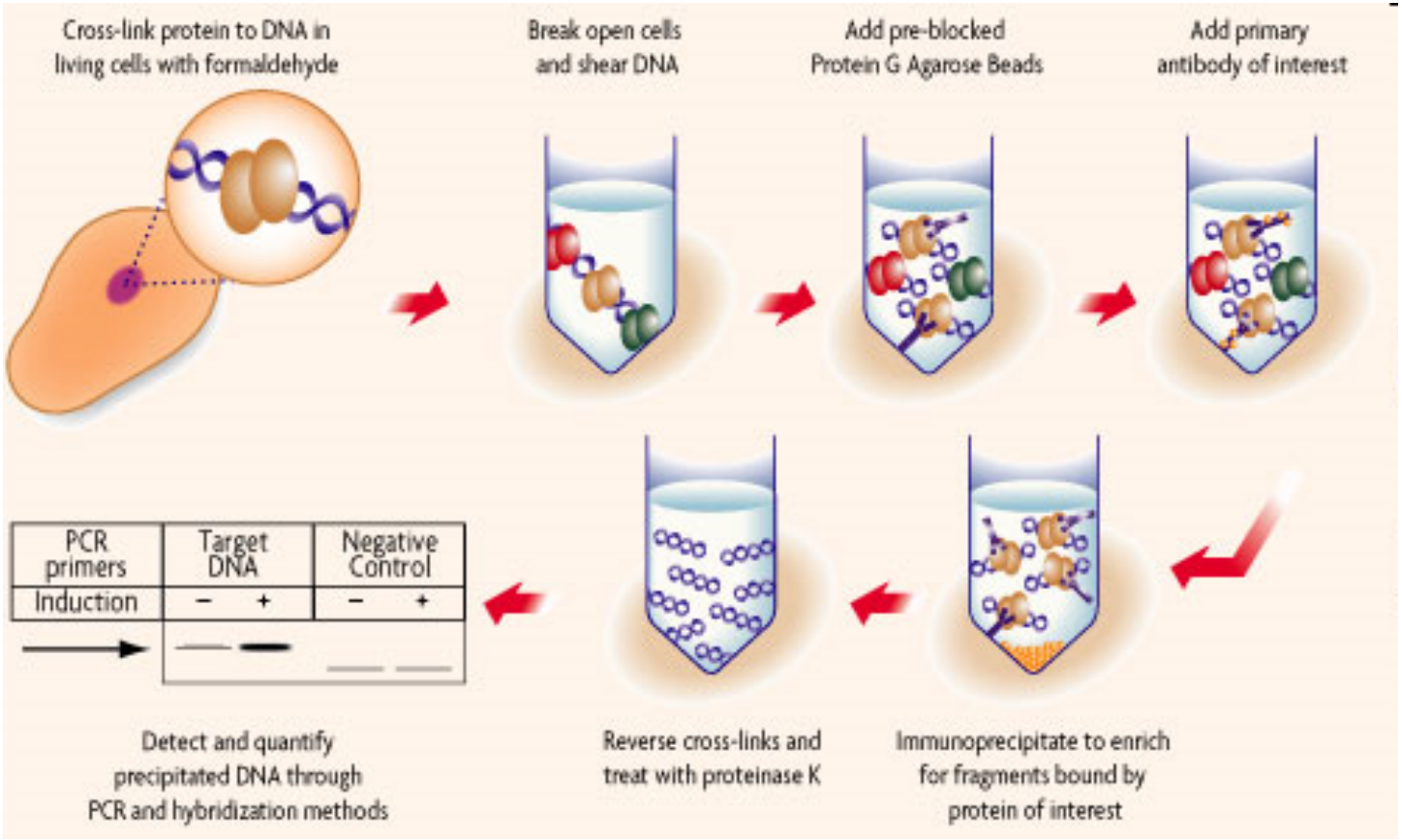
2



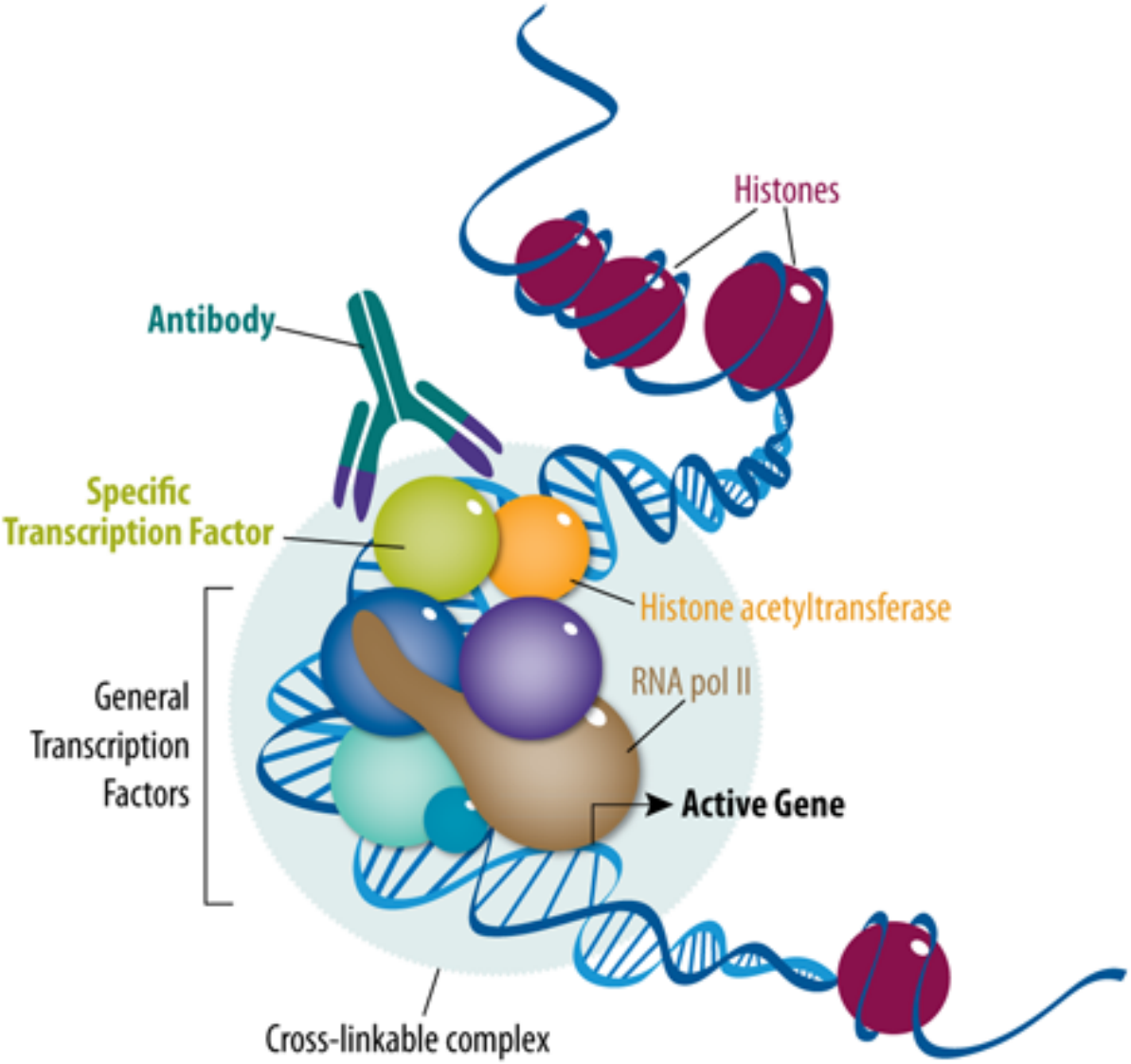
3



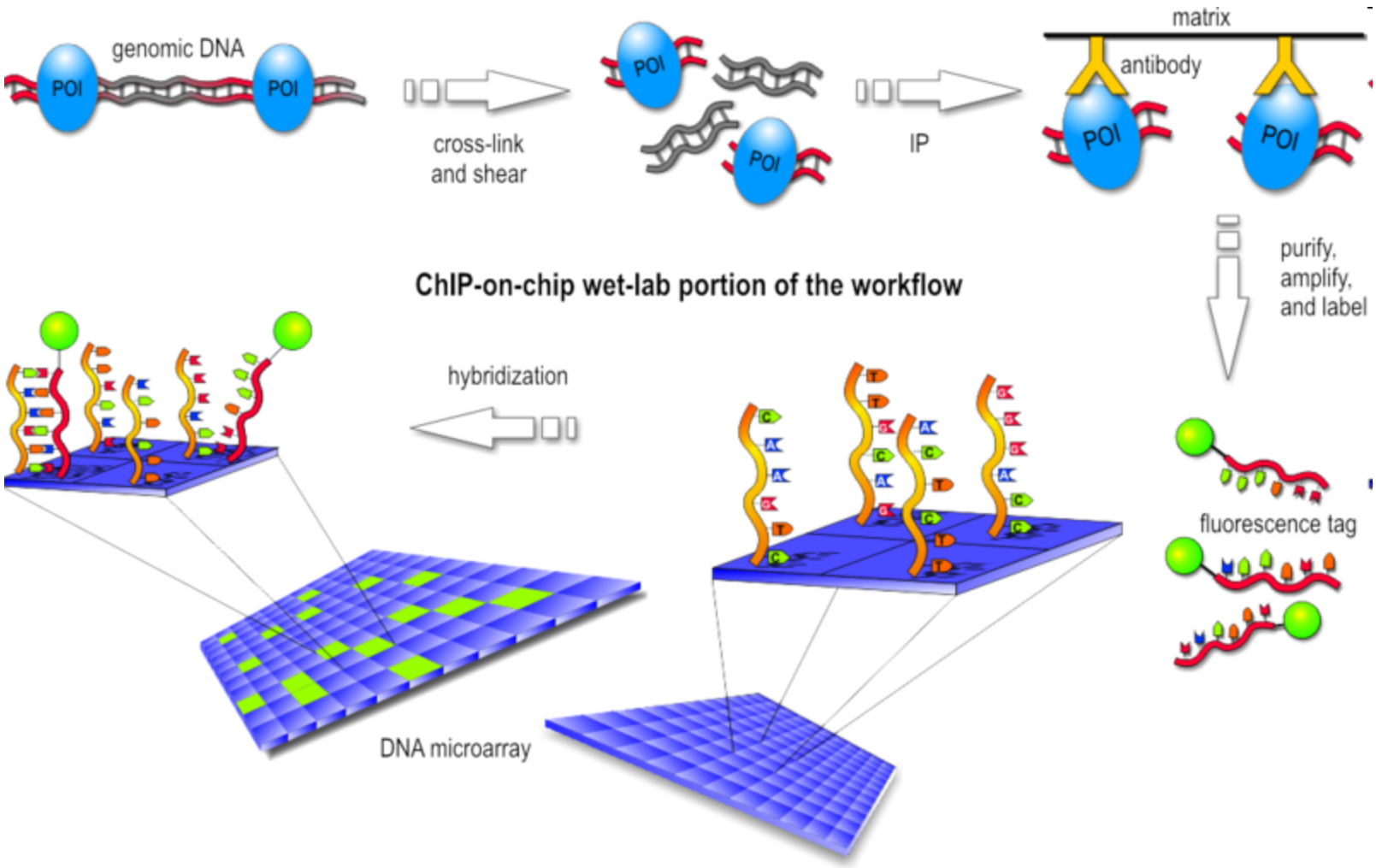
Chromatin immunoprecipitation (ChIP)



What do you get from the IP?



Protein-DNA interactions by ChIP-chip



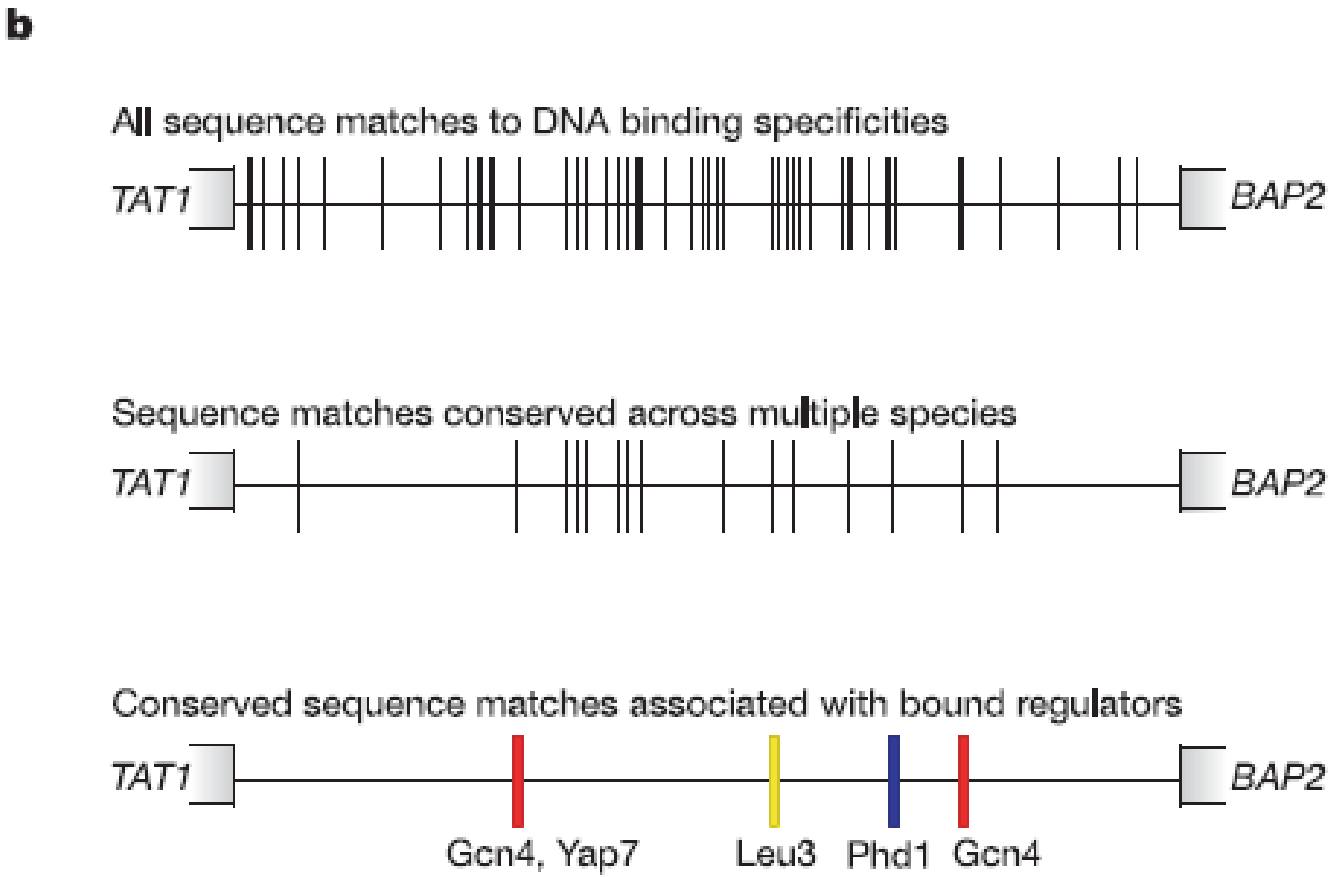
Transcriptional regulatory code of a eukaryotic genome

Christopher T. Harbison[1,2*], D. Benjamin Gordon[1*], Tong Ihn Lee[1], Nicola J. Rinaldi[1,2], Kenzie D. Macisaac[3], Timothy W. Danford[3], Nancy M. Hannett[1], Jean-Bosco Tagne[1], David B. Reynolds[1], Jane Yoo[1], Ezra G. Jennings[1], Julia Zeitlinger[1], Dmitry K. Pokholok[1], Manolis Kellis[1,3,4], P. Alex Rolfe[3], Ken T. Takusagawa[3], Eric S. Lander[1,2,4], David K. Gifford[3,4], Ernest Fraenkel[1,3] & Richard A. Young[1,2,4]

letters to nature

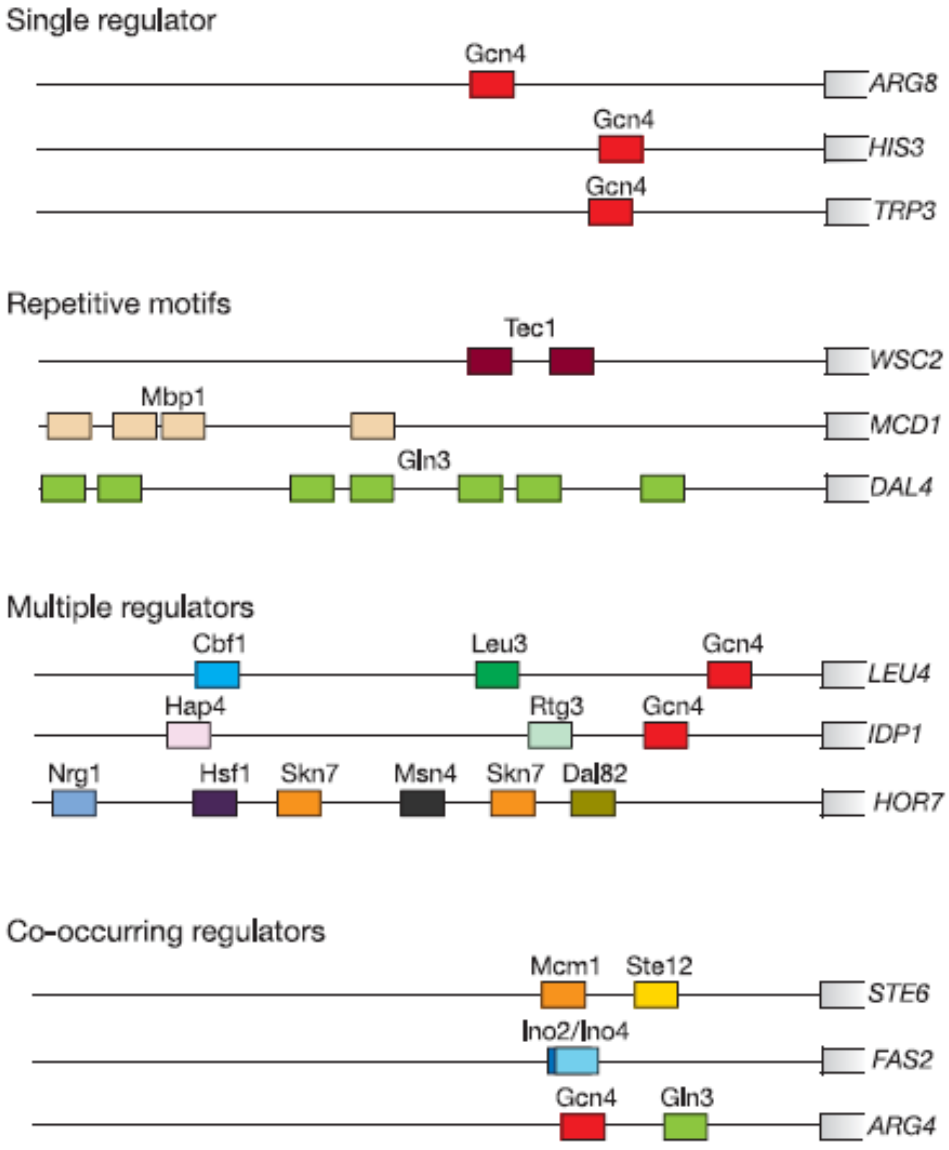
- Profiled 204 transcription factors in "normal" conditions
- Another 148 TFs in 13 additional growth perturbing conditions

Predicting the DNA binding sites



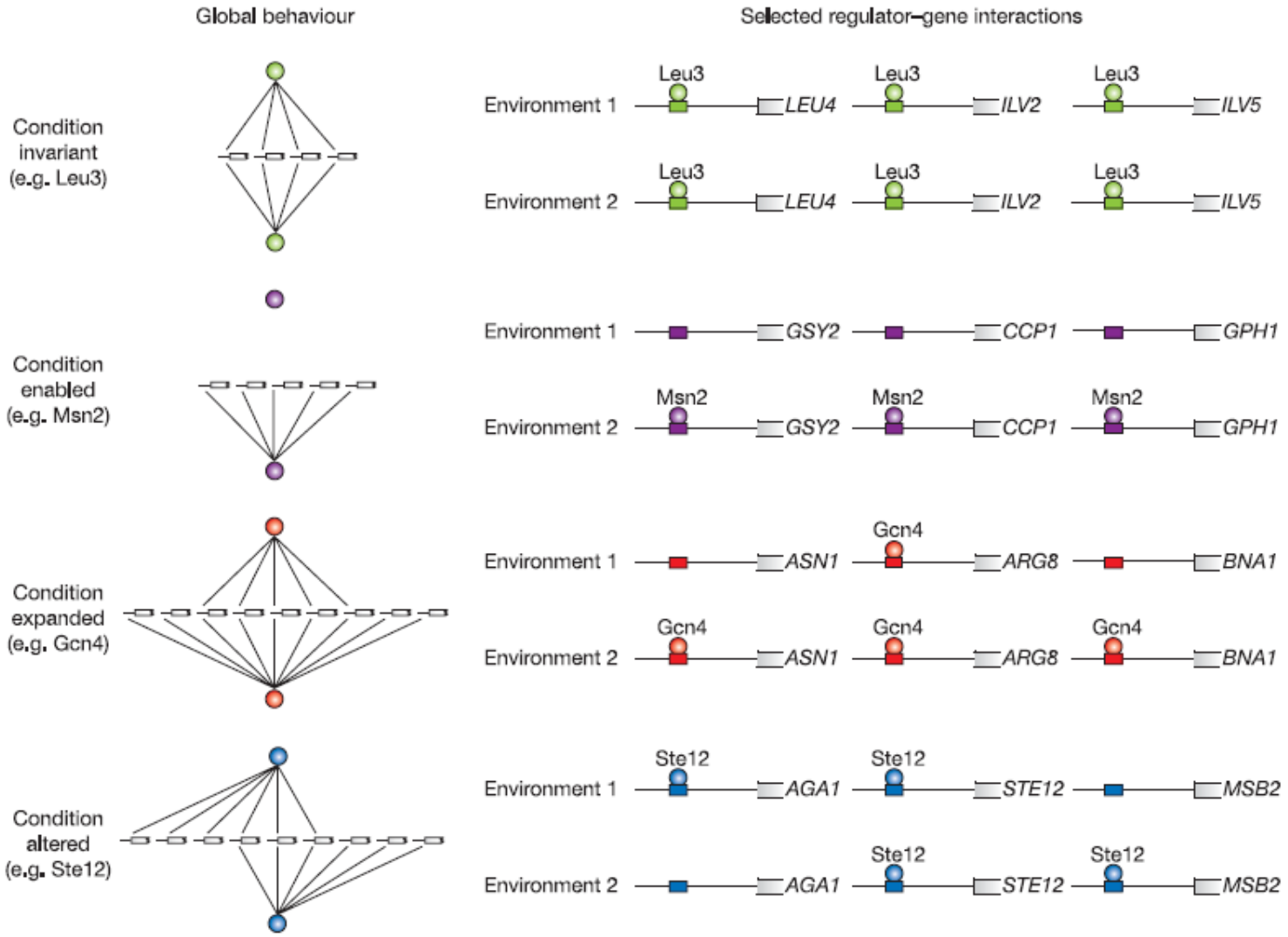
Harbison C, Gordon B, et al. Nature 2004

Different promoter architectures



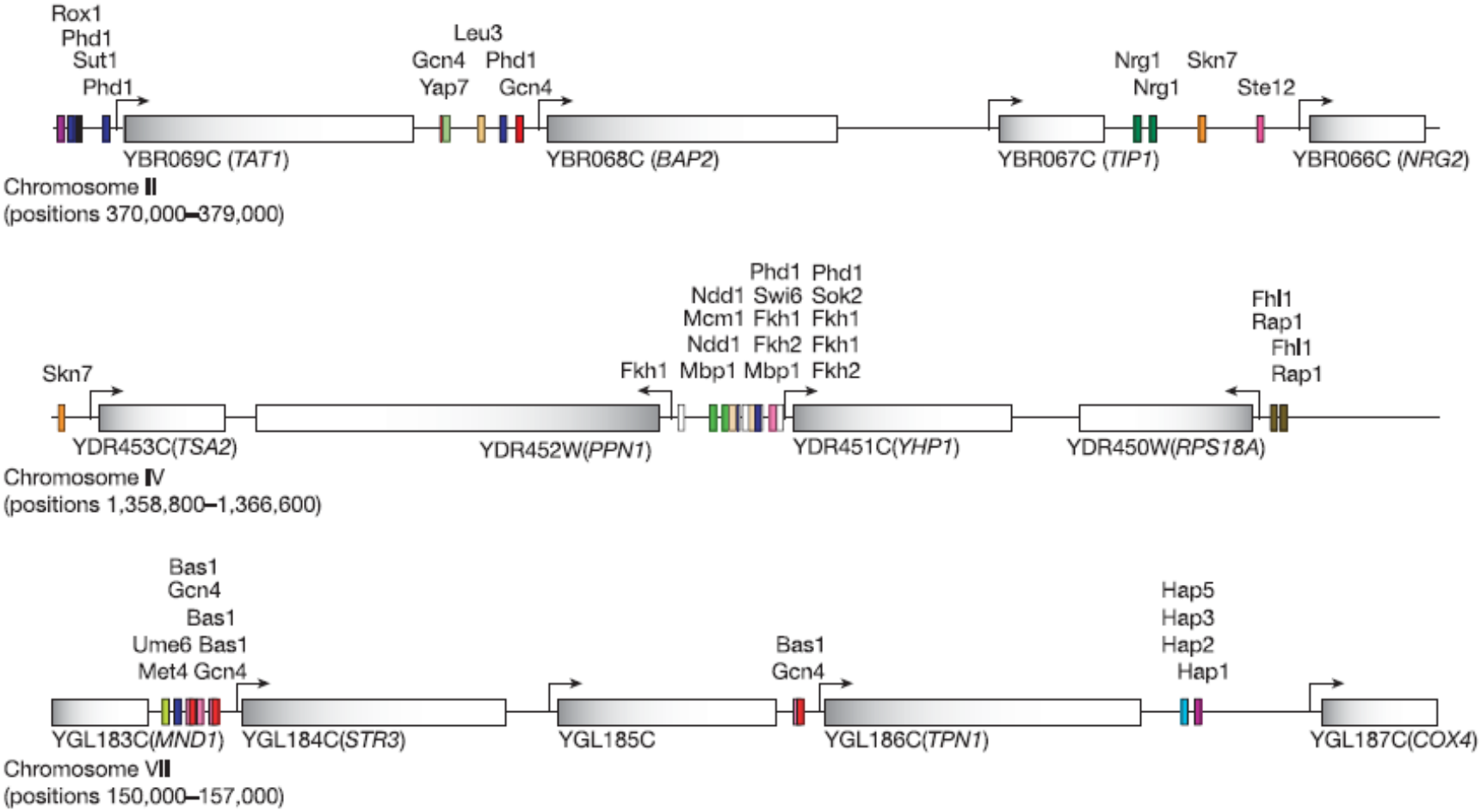
Harbison C, Gordon B, et al. Nature 2004

Different conditions activate TFs



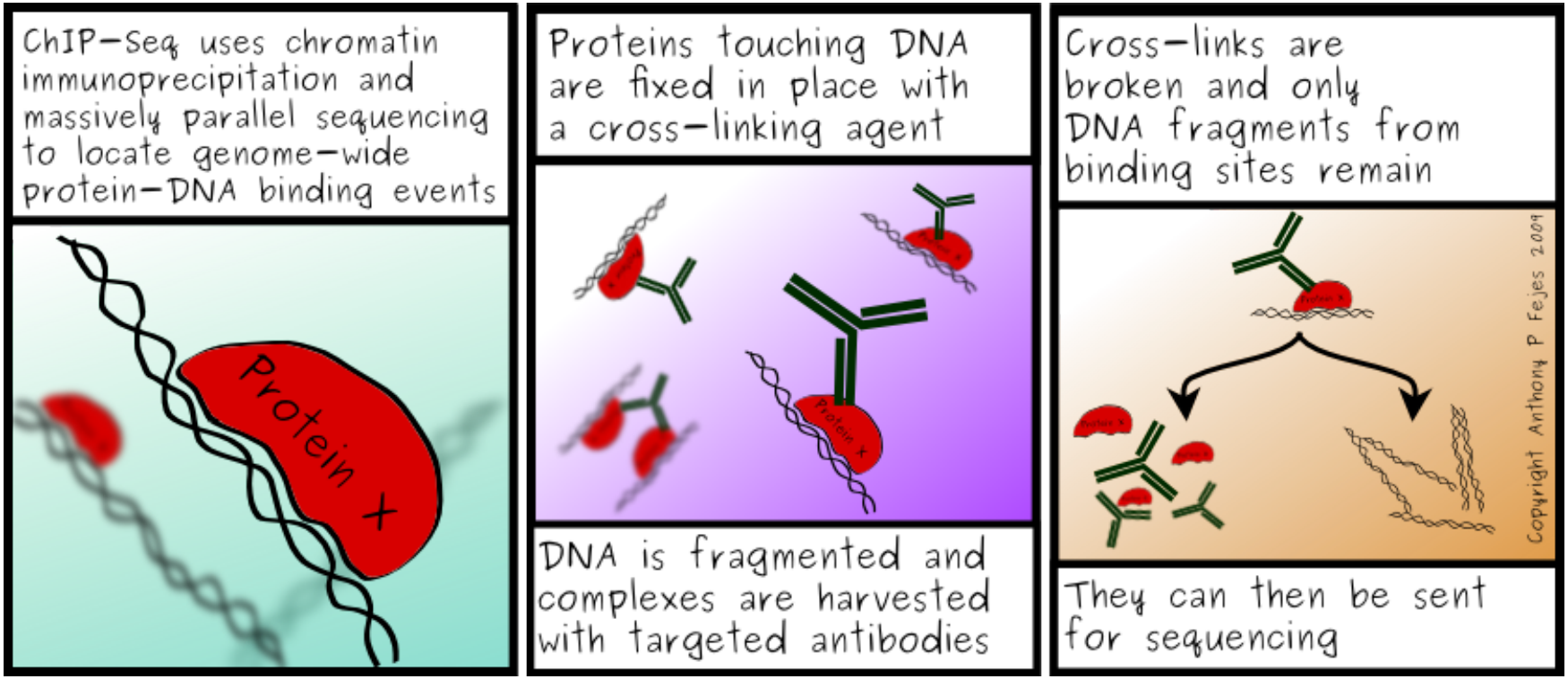
Harbison C, Gordon B, et al. Nature 2004

Example genome annotations based on chIP-chip

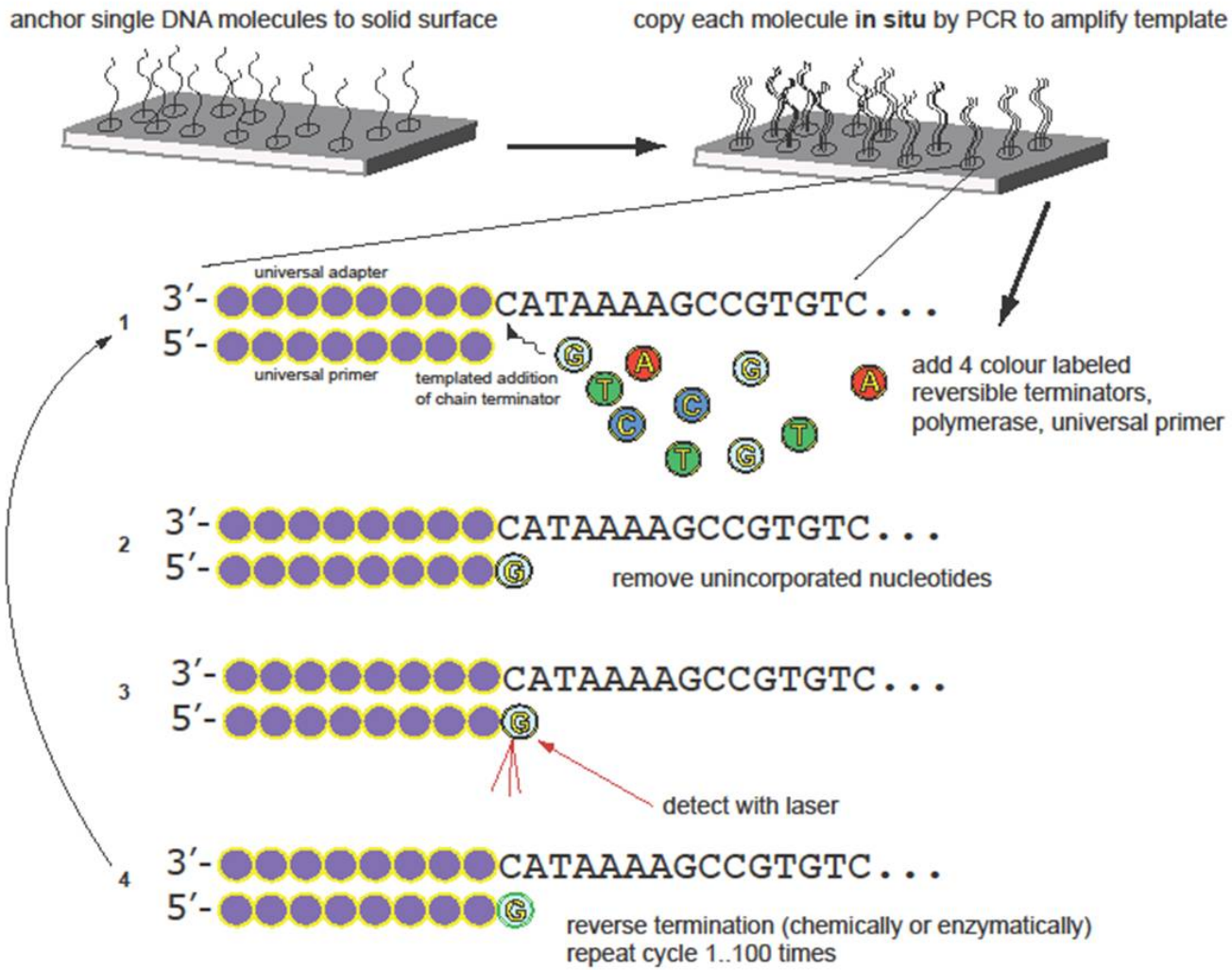


Harbison C, Gordon B, et al. Nature 2004

ChIP-Seq

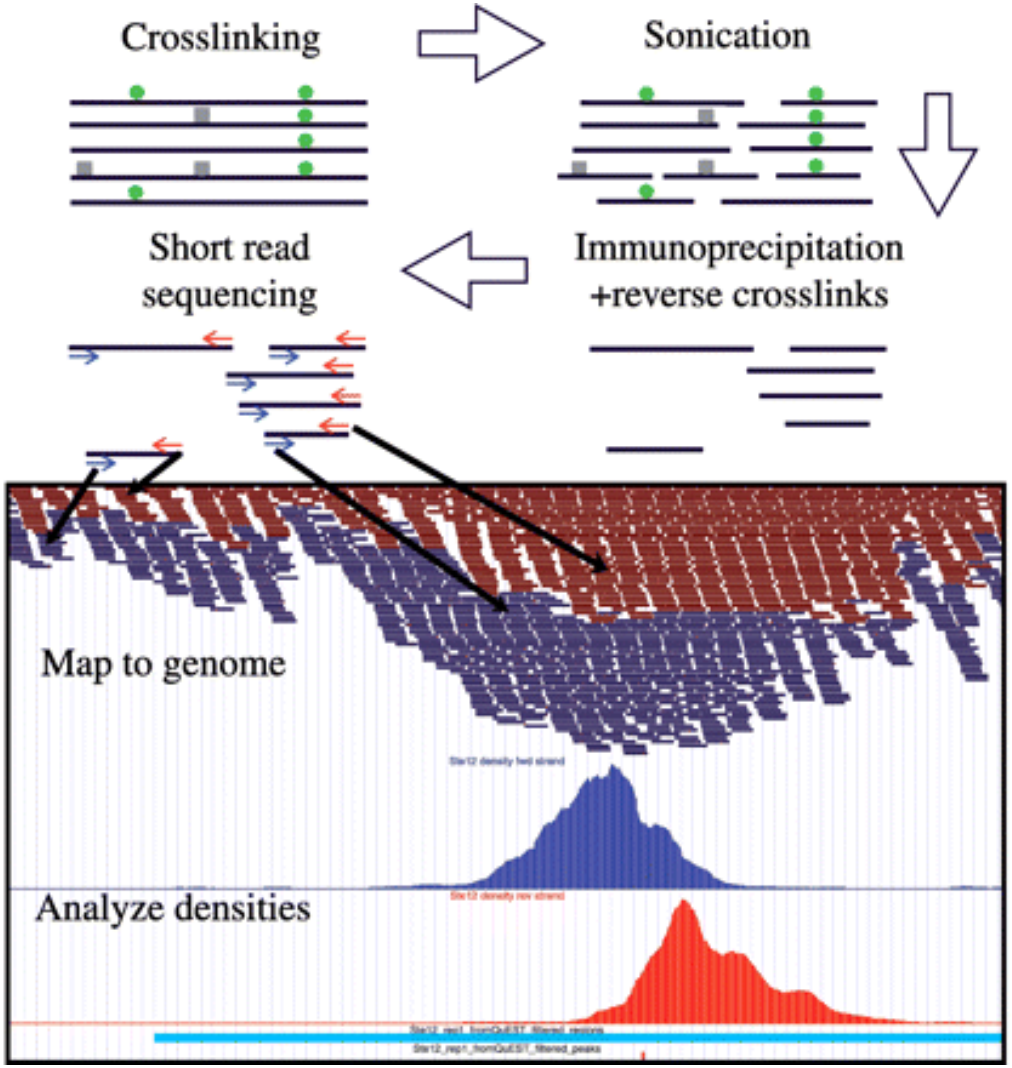

ChIP-Seq uses chromatin immunoprecipitation and massively parallel sequencing to locate genome-wide protein-DNA binding events
Protein X
Proteins touching DNA are fixed in place with a cross-linking agent
DNA is fragmented and complexes are harvested with targeted antibodies
Cross-links are broken and only DNA fragments from binding sites remain
Copyright Anthony P Fejes 2009
They can then be sent for sequencing

Next Generation Sequencing

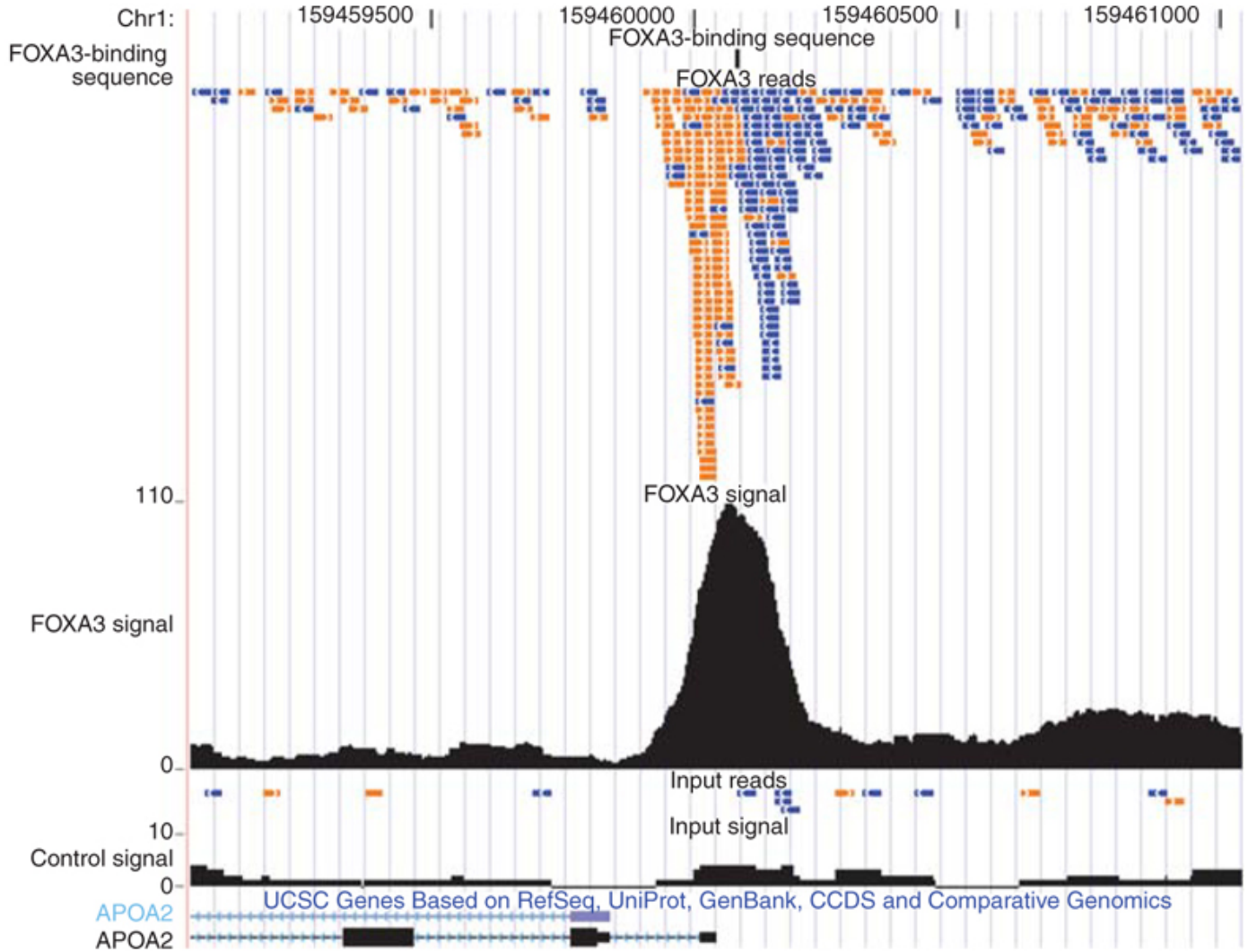


Let's see this in action

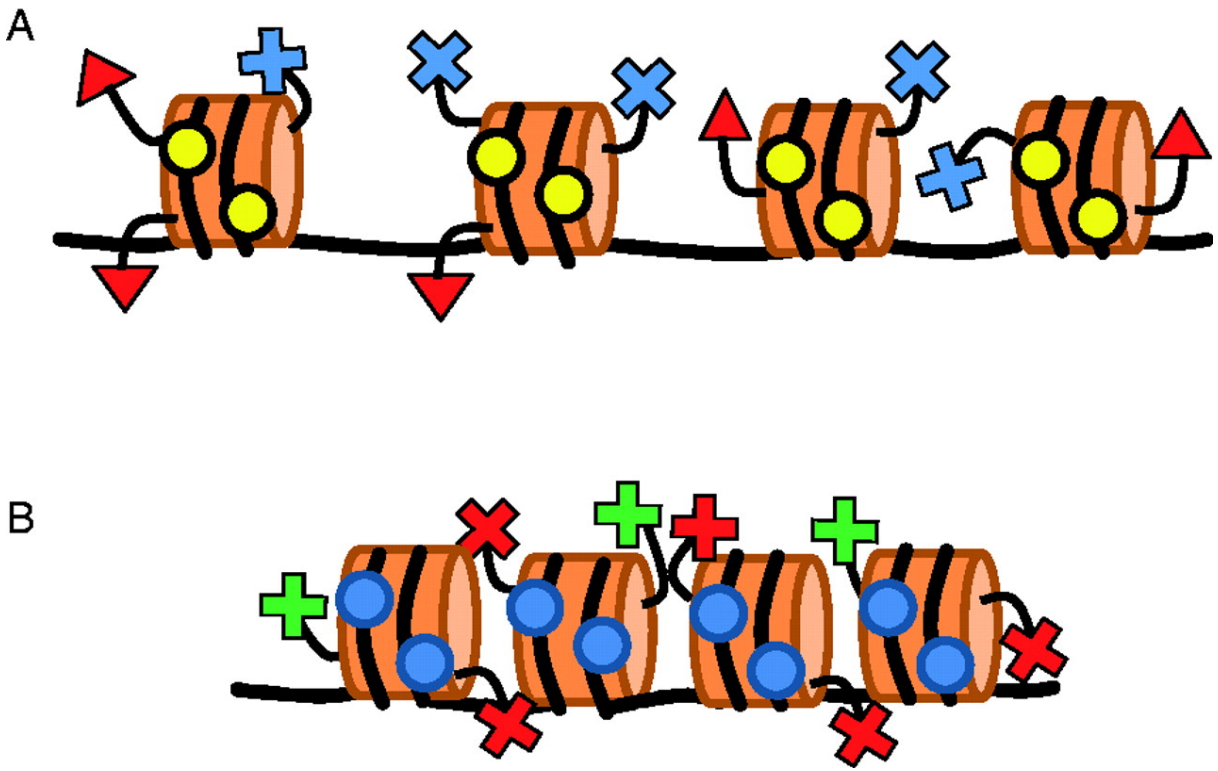
ChIP-Seq: Workflow is similar as for ChIP-chip



ChIP-Seq: the data is more precise



ChIP-Seq: Histone modifications



Emes R D , Farrell W E J Mol Endocrinol 2012;49:R19-R27

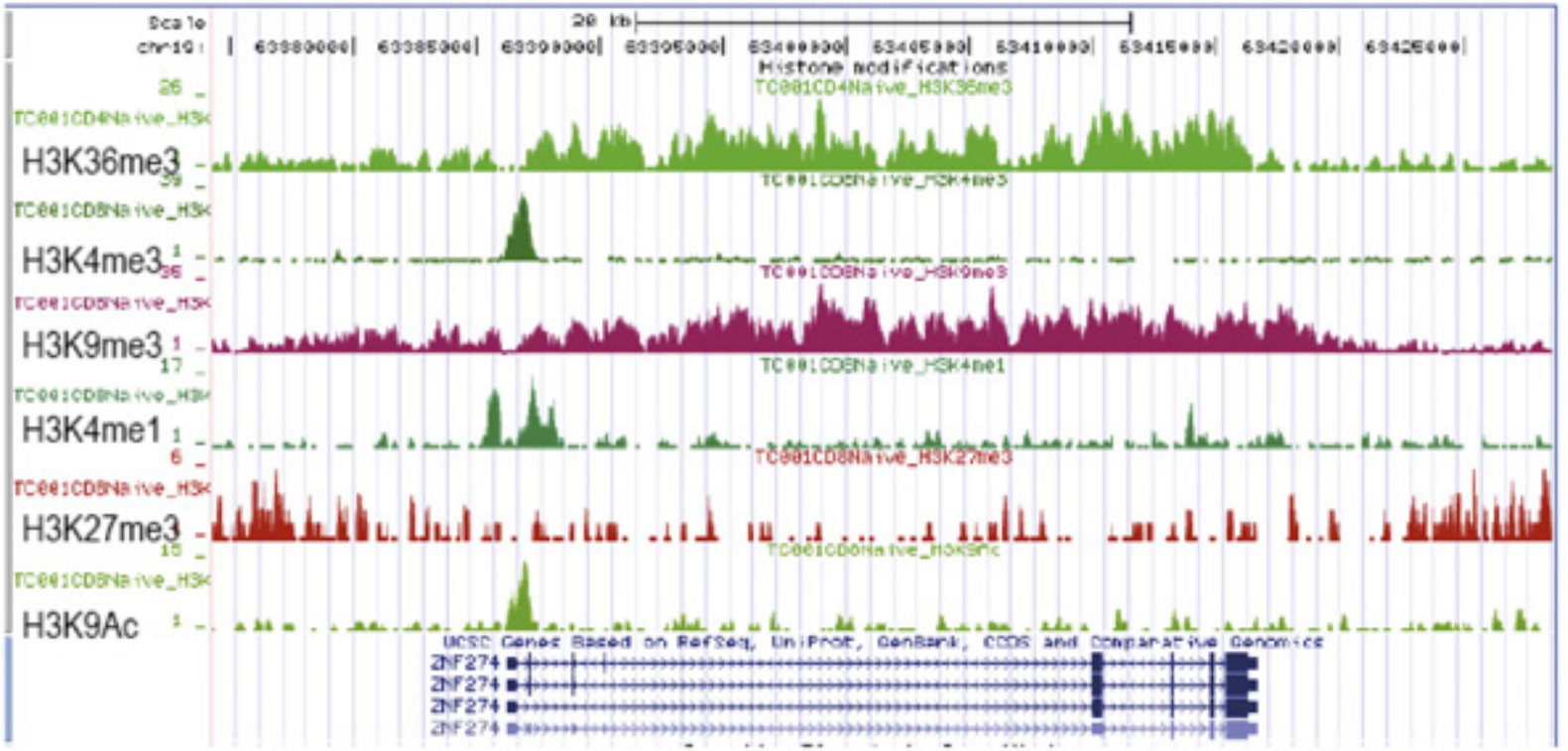
Published by BioScientifica

ChIP-Seq: Histone modifications

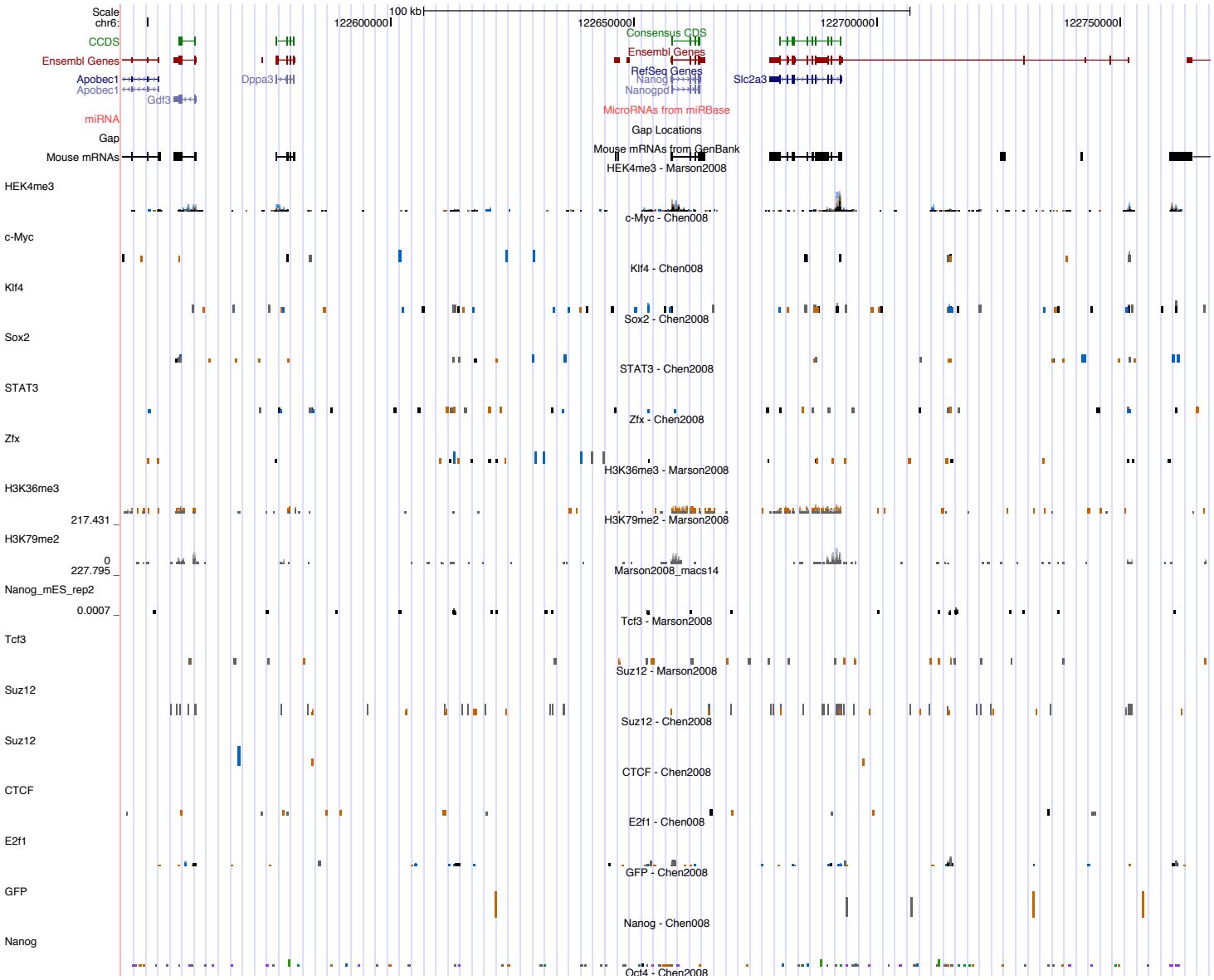
Type of modification	Histone						
	H3K4	H3K9	H3K14	H3K27	H3K79	H4K20	H2BK5
mono-methylation	activation	activation		activation	activation	activation	activation
di-methylation		repression		repression	activation		
tri-methylation	activation	repression		repression	activation, repression		repressior
acetylation		activation	activation				

- H3K4me3 is found in actively transcribed promoters, particularly just after the transcription start site.
- H3K9me3 is found in constitutively repressed genes.
- H3K27me is found in facultatively repressed genes.
- H3K36me3 is found in actively transcribed gene bodies.
- H3K9ac is found in actively transcribed promoters.
- H3K14ac is found in actively transcribed promoters.

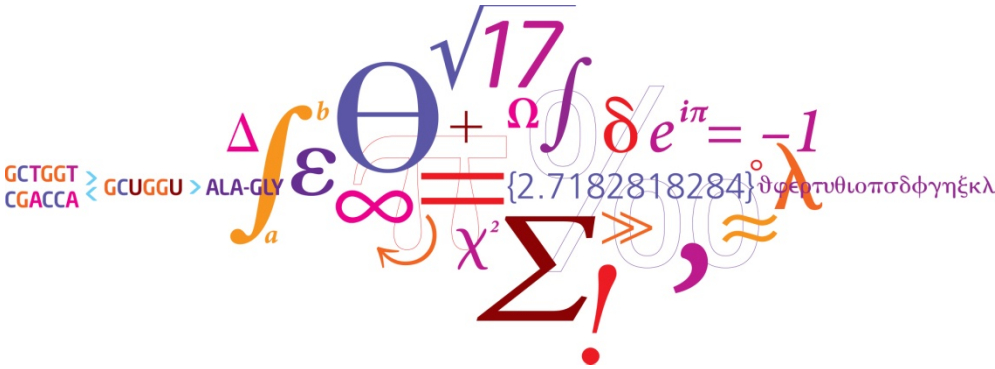
ChIP-Seq: Histone modifications



ChIP-Seq: Histone modifications

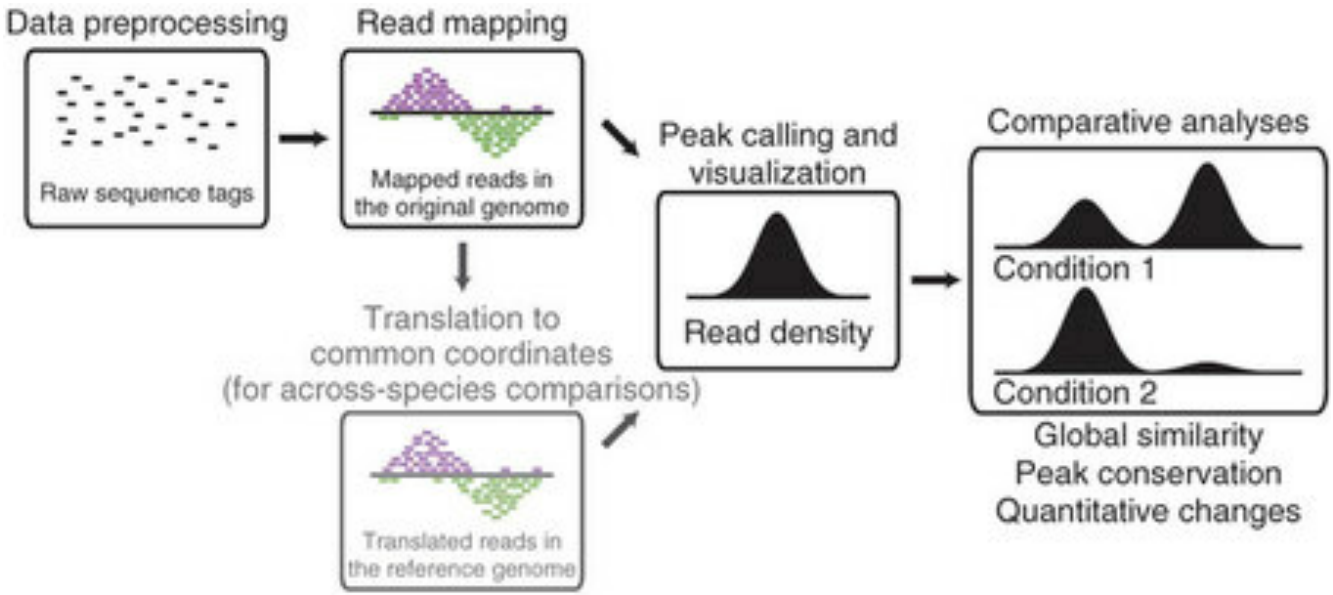


Analysing ChIP-Seq data

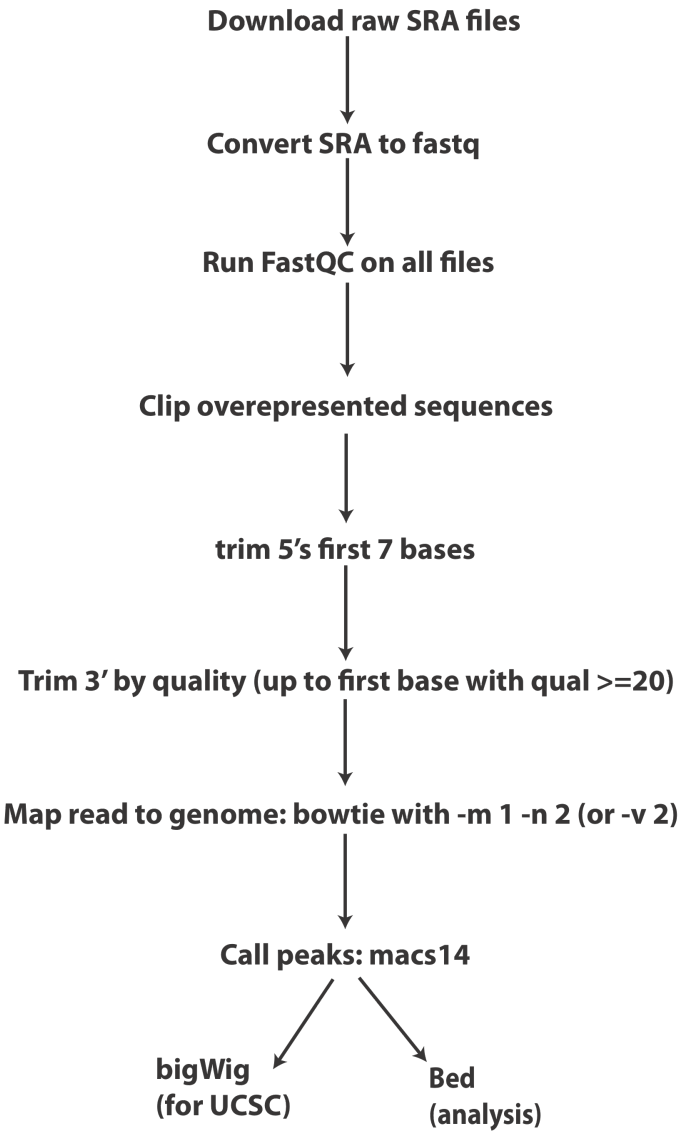


Center for Biological Sequence Analysis
Department of Systems Biology

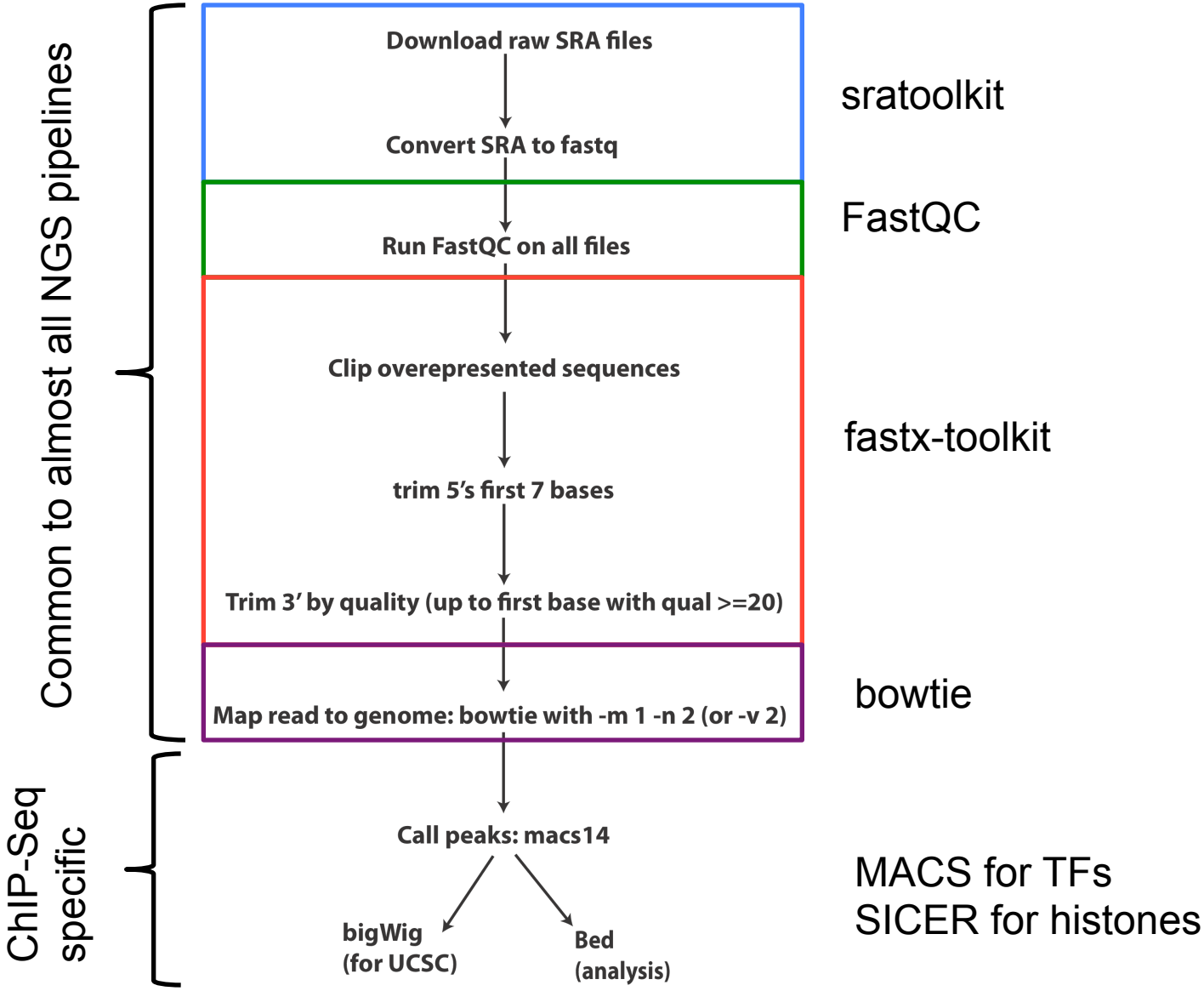
Analysis pipeline



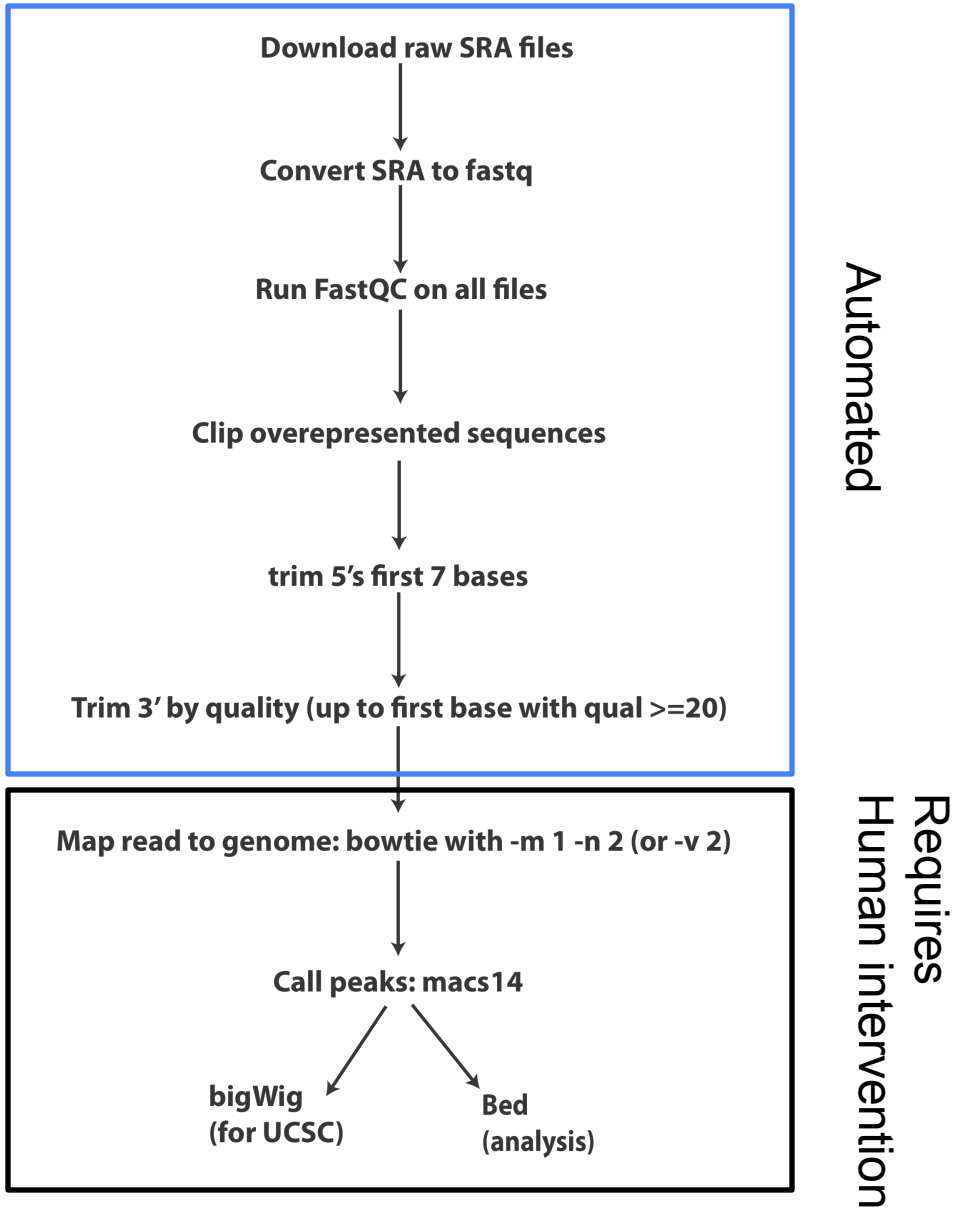
Analysis pipeline



Analysis pipeline



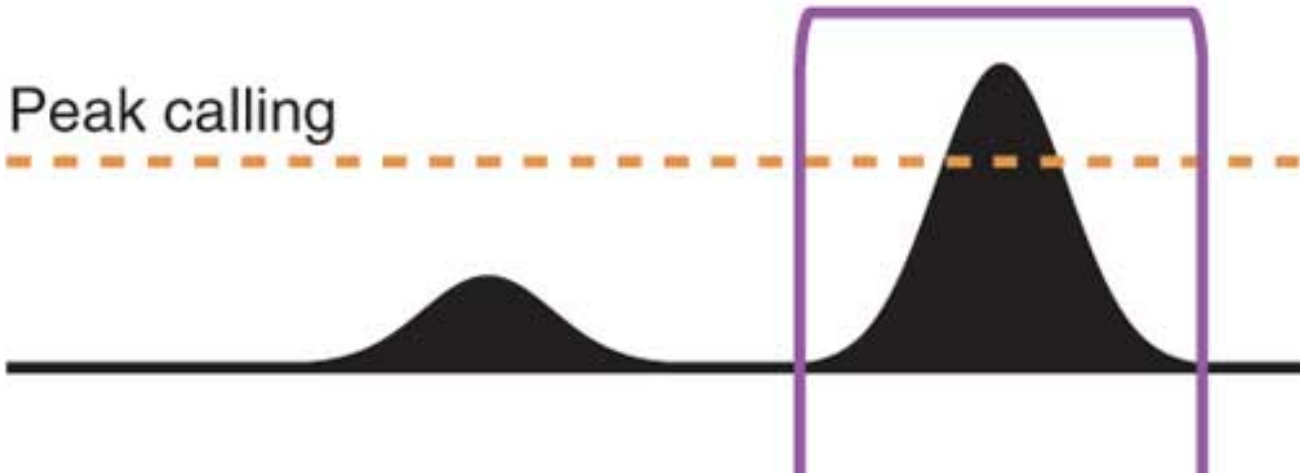
Analysis pipeline



Peak calling!

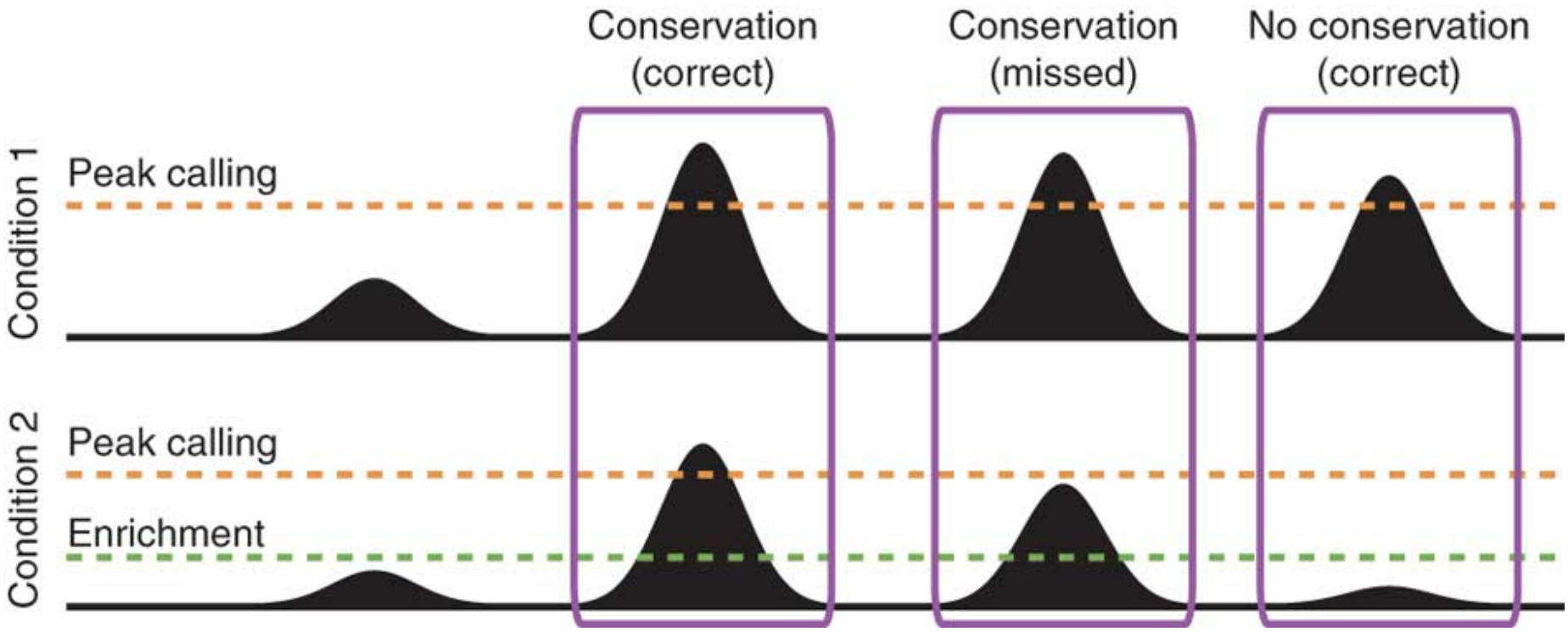
- Standard algorithm is MACS (Model-based Analysis of ChIP-Seq)
  - http://liulab.dfci.harvard.edu/MACS/index.html

Simple description:

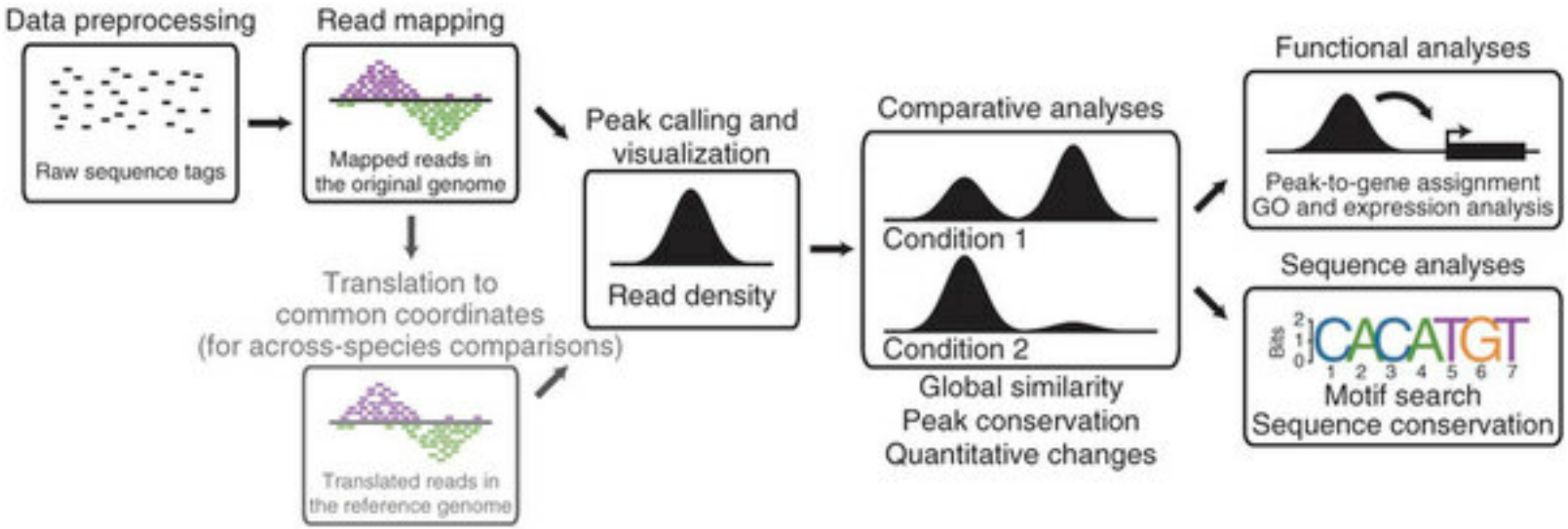


Peak calling!

– MACS allows for the calling of peaks using a background (control) track or against reactions performed in different conditions



Analysis pipeline



Downstream analysis: the biology!

- Having called the peaks, we need to find out what they are and how biologically relevant they are..
    - We need to find the DNA motifs bound by the TF
    - We need to call the peaks: what genes might be regulated by the TF?
    - We need to compare / integrate various datasets to get the full story

- Many tools have been created for this, but I like Homer (Heinz S, Benner C, Spann N, Bertolino E et Mol Cell 2010 May 28;38(4):576-589.)

http://biowhat.ucsd.edu/homer/ngs/index.html